

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2002, 10:59:45 ; Search time 18.85 Seconds

(without alignments)  
63.677 Million cell updates/sec

Title: US-09-785-059-3

Perfect score: 176

Sequence: 1 RMIRVQRWCRAIRHMRIRROGLRRMLRVV 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	50.6	856	1 ENV_HV1H2	P04578 human immun
2	89	50.6	856	1 ENV_HV1L1	P070626 human immun
3	89	50.6	861	1 ENV_HV1B1	P03377 human immun
4	88	50.0	847	1 ENV_HV1W2	P05880 human immun
5	88	50.0	856	1 ENV_HV1W1	P18792 human immun
6	82	46.6	846	1 ENV_HV1D1	P18799 human immun
7	80	44.5	856	1 ENV_HV1H3	P04624 human immun
8	79	44.9	852	1 ENV_HV1B1	P12488 human immun
9	79	44.9	853	1 ENV_HV1E1	P04581 human immun
10	78	44.3	853	1 ENV_HV1F1	P19531 human immun
11	78	44.3	853	1 ENV_HV1J2	P12487 human immun
12	78	44.3	856	1 ENV_HV1B1	P03375 human immun
13	78	44.3	856	1 ENV_HV1P1	P03378 human immun
14	77	43.8	851	1 ENV_HV1A2	P04582 human immun
15	76	43.2	851	1 ENV_HV1B8	P04582 human immun
16	72	40.9	848	1 ENV_HV1J1	P20871 human immun
17	71	40.3	865	1 ENV_HV1R4	P04579 human immun
18	70	39.8	868	1 ENV_HV1C4	P05879 human immun
19	69	39.2	859	1 ENV_HV1M1	P04583 human immun
20	68	38.6	856	1 ENV_HV1M1	P05877 human immun
21	68	38.6	863	1 ENV_HV1Z8	P05882 human immun
22	68	38.6	867	1 ENV_HV1J3	P12489 human immun
23	67	38.1	852	1 ENV_HV1S3	P19549 human immun
24	67	38.1	856	1 ENV_HV1S3	P05878 human immun
25	66	37.5	855	1 ENV_HV1OY	P20888 human immun
26	66	37.5	856	1 ENV_HV1Z8	P05881 human immun
27	65	36.9	843	1 ENV_HV1Y2	P25961 human immun
28	65	36.9	847	1 ENV_HV1S1	P19550 human immun
29	64	36.4	861	1 ENV_HV1K1	P18189 human immun
30	60	34.1	855	1 ENV_HV1Z6	P04580 human immun
31	58	33.0	854	1 ENV_SIVC2	P17281 chimpanzee
32	56.5	32.1	491	1 DGT1_CERAE	O69m71 ceropithec
33	53.5	30.4	488	1 DGT1_HUMAN	O75907 homo sapien

34	52	29.5	1459	1 YFIM_CAREL	O21874 caenorhabd1
35	51.5	29.3	498	1 DGT1_MOUSE	O92247 mus musculu
36	51.5	29.3	498	1 DGT1_RAT	O92247 mus musculu
37	51	29.0	556	1 PDPR_MOUSE	O15530 mus musculu
38	51	29.0	559	1 PDPR_MOUSE	O92240 mus musculu
39	51	29.0	559	1 PDPR_MOUSE	O55173 rattus norv
40	50	28.4	330	1 OUTG_EMENT	P25416 emeritcella
41	49.5	28.1	993	1 RPTG_ARATH	O24600 arabidopsis
42	49	27.8	197	1 US31_HCMVA	P09707 human cytom
43	49	27.8	501	1 GLPD_ECOLI	P13035 escherichia
44	48.5	27.6	242	1 RPO_SCVIA	P23172 saccharomyc
45	48.5	27.6	424	1 IMPB_SALTY	P18642 salmonella

## ALIGNMENTS

RESULT 1	ENV_HV1H2	STANDARD;	PRT;	856 AA.
ID	ENV_HV1H2			
AC	P04578; 009779;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Envelope glycoprotein GP160 precursor [contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].			
GN	ENV.			
OS	Human immunodeficiency virus type 1 (HXB2 isolate) (HIV-1).			
OC	Viruses; Retroid viruses; Retroviridae; Lentiviruses.			
OX	NCBI_Taxid=11706;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87299196; PubMed=3040055;			
RA	Ratner L., Fisher A., Jagodzinski L.L., Mitsuya H., Liou R.-S.,			
RA	Gallo R.C., Wong-Staal F.;			
RT	"Complete nucleotide sequences of functional clones of the AIDS virus".			
RL	AIDS Res. Hum. Retroviruses 3:57-69(1987).			
RL	[2]			
RP	REVISIONS.			
RA	Ratner L., Fisher A., Jagodzinski L.L., Mitsuya H., Liou R.-S.,			
RA	Gallo R.C., Wong-Staal F.;			
RL	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; K03455; AAB50262.1; -			
DR	EMBL; AF038399; AAB99976.1; -			
DR	EMBL; AF038319; AAC82596.1; -			
DR	HIV; K03455; ENV\$HXB2.			
DR	InterPro; IPR000328; Env_GP41.			
DR	InterPro; IPR000777; GP120.			
DR	Pfam; PF00516; GP120; 1.			
DR	Pfam; PF00517; GP41; 1.			
KW	AIDS; Coat protein; Polypeptide; Glycoprotein; Transmembrane;			
KW	Signal.			
FT	SIGNAL	1	30	
FT	CHAIN	31	511	EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	512	856	TRANSMEMBRANE GLYCOPROTEIN.
FT	DISULFID	54	74	BY SIMILARITY.
FT	DISULFID	119	205	BY SIMILARITY.
FT	DISULFID	126	196	BY SIMILARITY.
FT	DISULFID	131	157	BY SIMILARITY.
FT	DISULFID	218	247	BY SIMILARITY.
FT	DISULFID	228	239	BY SIMILARITY.
FT	DISULFID	296	331	BY SIMILARITY.
FT	DISULFID	378	445	BY SIMILARITY.

```

FT DISULFID 385 418 BY SIMILARITY.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 750 750 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 856 AA; 97212 MW; 6FAB16AF85107E0 CRC64;

```

```

Query Match 50.6%; Score 89; DB 1; Length 856;
Best Local Similarity 75.0%; Pred. No. 8e-05;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Oy 1 RWIRVQRCRAIRHRIIRIGLRWL 28
Db 828 RVEIVVQACRAIRHRIIRIGLERIL 855

```

```

RESULT 2
ENV_HVILW STANDARD; PRT; 856 AA.
ID ENV_HVILW
AC 070626;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (HIV-1 isolate) (HIV-1).
OC Viruses; Retrovirus; Retroviridae; Lentivirus.
OX NCBI_TaxID=82834;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95127297; Pubmed=7826699;
RA Reitz M.S. Jr., Hall L., Robert-Guroff M., Lautenberger J., Hahn B.M.,
Shaw G.M., Kong L.I., Weiss S.H., Waters D., Gallo R.C., Blattner W.;
RT Viral variability and serum antibody response in a laboratory worker
RL AIDS Res. Hum. Retroviruses 10:1143-1155(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
CC EMBL: U12055; AAA6690.1; -.
CC GlycosultedB; 070626; -.
CC InterPro: IPR000328; Env_GP41.
CC InterPro: IPR000777; GP120.
CC Pfam: PF00516; GP120; 1.
CC DR Pfam: PF00517; GP41; 1.
CC KW Aids; Coat protein; Polypeptide; Glycoprotein; Transmembrane;
KW Signal.
FT CHAIN 1 30
FT SIGNAL 31 511
FT CHAIN 512 856
FT DISULFID 54 74
FT DISULFID 119 205
FT DISULFID 126 196
FT DISULFID 131 157
FT DISULFID 218 247
FT DISULFID 228 239
FT DISULFID 296 331
FT DISULFID 378 445
FT DISULFID 385 418
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 750 750 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 856 AA; 96938 MW; 0C241332CFE687 CRC64;

```

```

Query Match 50.6%; Score 89; DB 1; Length 856;
Best Local Similarity 75.0%; Pred. No. 8e-05;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Oy 1 RWIRVQRCRAIRHRIIRIGLRWL 28
Db 828 RVEIVVQACRAIRHRIIRIGLERIL 855

```

```

RESULT 3
ENV_HVILR STANDARD; PRT; 861 AA.
ID ENV_HVILR
AC P03377;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].

```

[illegible]

```

SQ SEQUENCE      861 AA;   97487 MW;    04DE2B4D4EFD63A CRC64;
Query Match          50.6%; Score 89; DB 1; Length 861;
Best Local Similarity 75.0%; Pred. No. 8e-05;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 RWIRVQWMCRAIRIMRRIRGRLRWL 28
        | | | | | | | | | | | | | | | |
DB      833 RVIEVGACRAIRHPRIRIGLERIL 860

RESULT      4
ENV_HY1W2 STANDARD: PRT; 847 AA.
AC P05880;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope polypeptide GP160 precursor [Contains: Exterior membrane
DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (WMJ2 isolate) (HTV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11705;
[ ]
RP SEQUENCE FROM N.A.
RX MEDLINE=86235450; PubMed=3012778;
RA Hahn B.H., Shaw G.M., Taylor M.E., Redfield R.R., Markham P.D.,
RA Salahuddin S.Z., Wong-staal F., Gallo R.C., Parks E.S., Parks W.P.;
R "Genetic variation in HTLV-III/LAV over time in patients with AIDS or
RT at risk for AIDS.";
RL -I- MISCELLANEOUS: ISOLATES WMJ1, WMJ2, AND WMJ3 WERE OBTAINED FROM
CC BLOOD SAMPLES SEQUENTIALLY TAKEN FROM A TWO-YEAR OLD HAITIAN WHO
CC WAS PERINATALLY INFECTED BY HER MOTHER.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M12507; AAB12990.1; -.
DR DR HIT: M12507; ENVSMWJ2.
DR InterPro: IPR000328; Env_GP41.
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW AIDS; Coat protein; Polypeptide; Glycoprotein; Transmembrane;
KW signal.
FT SIGNAL. 1 29
FT CHAIN 30 501 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 502 847 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 202 BY SIMILARITY.
FT DISULFID 125 193 BY SIMILARITY.
FT DISULFID 130 152 BY SIMILARITY.
FT DISULFID 215 244 BY SIMILARITY.
FT DISULFID 225 326 BY SIMILARITY.
FT DISULFID 293 326 BY SIMILARITY.
FT DISULFID 372 435 BY SIMILARITY.
FT DISULFID 379 408 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

FT	DISULFID	296	330	BY SIMILARITY.
FT	DISULFID	376	444	BY SIMILARITY.
FT	DISULFID	383	417	BY SIMILARITY.
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	140	140	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	151	151	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	155	155	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	197	197	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	241	241	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	262	262	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	276	276	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	289	289	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	295	295	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	338	338	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	354	354	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	390	390	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	394	394	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	447	447	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	459	459	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	625	625	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	637	637	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	856 AA;	97526 MW;	DB68BD1E49C404DE9 CRC64;

```

Query Match          50.0%; Score 88; DB 1; Length 856;
Best Local Similarity 75.0%; Pred. NO. 0.00011;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1  EWIRVVGRCRAIRHWRIRRIQGLRWL 28
      1  | | | | | | | | | | | | | | | |
Db      828  RVEIVVGRCIRALIHIPRIIRQGLRWL 855
      1  | | | | | | | | | | | | | | | |

RESULT 6
ENV_HVIND  STANDARD;  PRT;  846 AA.
ID  ENV_HVIND
AC  P18799;
DT  01-NOV-1990 (Rel. 16, Created)
DT  01-NOV-1990 (Rel. 16, Last sequence update)
DT  13-JUL-1999 (Rel. 38, Last annotation update)
DE  Envelope polypeptide GP160 precursor [Contains: Exterior membrane
DE  glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
DE  ENV.
OS  Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).
OS  Viruses; Retroid viruses; Retroviridae; Lentivirus.
OC  NCBI_TaxID=11695;
OX  [1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=90034200; Pubmed=2806917;
RX  Spire B., Sire J., Zachar V., Rey F., Barre-Sinoussi F., Gilbert F.,
RA  Hampe A., Chermann J.C.;
RA  "Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the
RT  human immunodeficiency virus.";
RT  Gene 81:275-284(1989).
RL  -I- MISCELLANEOUS: (NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH
CC  AIDS, AND IS A HIGHLY CYTOPATHOGENIC STRAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----

```

DR EMBL: M27323; AAA44873.1; -  
DR PIR: J00066; VCLIND.  
DR HIV: M27323; ENV5NDK.  
DR InterPro: IPR000328; Env\_GP41.  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120; 1.  
DR Pfam: PF00517; GP41; 1.  
KW Aids; Coat protein; Polypeptide; Glycoprotein; Transmembrane;  
Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 501 EXTERIOR MEMBRANE GLYCOPROTEIN.  
FT CHAIN 502 846 TRANSMEMBRANE GLYCOPROTEIN.  
FT DISULFID 53 73 BY SIMILARITY.  
FT DISULFID 118 200 BY SIMILARITY.  
FT DISULFID 125 191 BY SIMILARITY.  
FT DISULFID 130 152 BY SIMILARITY.  
FT DISULFID 213 242 BY SIMILARITY.  
FT DISULFID 223 234 BY SIMILARITY.  
FT DISULFID 291 328 BY SIMILARITY.  
FT DISULFID 374 435 BY SIMILARITY.  
FT DISULFID 381 408 BY SIMILARITY.  
FT CARBOHYD 87 408 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 438 438 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 846 AA; 96476 MW; 8A3B9DA527DE2E83 CRC64;  
Query Match 46.6%; Score 82; DB 1; Length 846;  
Best Local Similarity 67.9%; Pred. No. 0.00072;  
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

RA Crowl R., Ganguly K., Gordon M., Conroy R., Schaber M., Kramer R.,  
RT Shaw G.M., Wong-Staal F., Reddy E.P.;  
RA "HTLV-III env gene products synthesized in E. coli are recognized by  
RT antibodies present in the sera of AIDS patients.";  
RL Cell 41:979-986(1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M14100; AAA44679.1; -  
DR HIV: M14100; ENV5HXB3.  
DR InterPro: IPR000328; Env\_GP41.  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120; 1.  
DR Pfam: PF00517; GP41; 1.  
KW Aids; Coat protein; Polypeptide; Glycoprotein; Transmembrane;  
Signal.  
FT SIGNAL 1 30  
FT CHAIN 31 511 EXTERIOR MEMBRANE GLYCOPROTEIN.  
FT CHAIN 512 856 TRANSMEMBRANE GLYCOPROTEIN.  
FT DISULFID 119 205 BY SIMILARITY.  
FT DISULFID 126 196 BY SIMILARITY.  
FT DISULFID 131 157 BY SIMILARITY.  
FT DISULFID 218 247 BY SIMILARITY.  
FT DISULFID 228 239 BY SIMILARITY.  
FT DISULFID 286 331 BY SIMILARITY.  
FT DISULFID 378 445 BY SIMILARITY.  
FT DISULFID 385 418 BY SIMILARITY.  
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 750 750 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 856 AA; 97188 MW; 3373C68BB84C1AFC CRC64;  
Query Match 45.5%; Score 80; DB 1; Length 856;  
Best Local Similarity 71.4%; Pred. No. 0.0014;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

DB      828 RVEVVOEAYRAIRHPRIRROGLERIL 855
RESULT  8
ENV_HV1BN STANDARD: PRT; 852 AA.
AC      P12488;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Envelope polypeptide GP160 precursor [Contains: Exterior membrane
DE      glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN      ENV.
OS      Human immunodeficiency virus type 1 (BR isolate) (HIV-1).
OC      Viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11693;
RN      [1]
RP      SEQUENCE FROM N.A. MEDLINE=2789516;
RX      MEDLINE=89085613; PubMed=2789516;
RA      Anand R., Thayer R., Srinivasan A., Nayyar S., Gardner M., Luciw P.,
RA      Danekar S.;
RT      "Biological and molecular characterization of human immunodeficiency
RT      virus (HIV-1BR) from the brain of a patient with progressive
RT      dementia.";
RL      Virology 168:79-89(1989).
CC      -1- MISCELLANEOUS: THIS VIRUS IS CYTOPATHICALLY ACTIVE AND WAS
CC      HARVESTED FROM THE BRAIN TISSUE OF A NEUROLOGICAL AIDS PATIENT.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M21098; AAA44221.1; -
DR      PIR; A31667; VCLJBR.
DR      HIV; M21098; ENVSBRAVA.
DR      InterPro: IPR000328; Env GP41.
DR      InterPro: IPR000777; GP120.
DR      Pfam; PF00516; GP120; 1.
DR      Pfam; PF00517; GP41; 1.
KW      Aids; Coat protein; Glycoprotein; Transmembrane;
KW      signal.
FT      SIGNAL 1 30
FT      CHAIN 31 507
FT      CHAIN 508 852
FT      DISULFID 54 74
FT      DISULFID 119 205
FT      DISULFID 126 196
FT      DISULFID 131 155
FT      DISULFID 218 247
FT      DISULFID 228 239
FT      DISULFID 296 330
FT      DISULFID 376 439
FT      DISULFID 383 412
FT      DISULFID 49 49
FT      CARBOHYD 88 88
FT      CARBOHYD 135 135
FT      CARBOHYD 138 138
FT      CARBOHYD 154 154
FT      CARBOHYD 158 158
FT      CARBOHYD 197 197
FT      CARBOHYD 234 234
FT      CARBOHYD 241 241
FT      CARBOHYD 262 262
FT      CARBOHYD 276 276
FT      CARBOHYD 289 289
FT      CARBOHYD 295 295
FT      CARBOHYD 301 301
FT      CARBOHYD 331 331
FT      CARBOHYD 354 354
FT      SIGNAL 31 30
FT      CHAIN 31 507
FT      CHAIN 508 852
FT      DISULFID 54 74
FT      DISULFID 119 205
FT      DISULFID 126 196
FT      DISULFID 131 155
FT      DISULFID 218 247
FT      DISULFID 228 239
FT      DISULFID 296 330
FT      DISULFID 376 439
FT      DISULFID 383 412
FT      DISULFID 49 49
FT      CARBOHYD 88 88
FT      CARBOHYD 135 135
FT      CARBOHYD 138 138
FT      CARBOHYD 154 154
FT      CARBOHYD 158 158
FT      CARBOHYD 197 197
FT      CARBOHYD 234 234
FT      CARBOHYD 241 241
FT      CARBOHYD 262 262
FT      CARBOHYD 276 276
FT      CARBOHYD 289 289
FT      CARBOHYD 295 295
FT      CARBOHYD 301 301
FT      CARBOHYD 331 331
FT      CARBOHYD 354 354

```

```

FT      CARBOHYD 360 360
FT      CARBOHYD 384 384
FT      CARBOHYD 390 390
FT      CARBOHYD 396 396
FT      CARBOHYD 400 400
FT      CARBOHYD 442 442
FT      CARBOHYD 456 456
FT      CARBOHYD 607 607
FT      CARBOHYD 612 612
FT      CARBOHYD 621 621
FT      CARBOHYD 633 633
FT      CARBOHYD 670 670
FT      CARBOHYD 812 812
SQ      SEQUENCE 852 AA; 97203 MW; 2BB866345DEC915F CRC64;
Query Match 44.9%; Score 79; DB 1; Length 852;
Best Local Similarity 69.0%; Pred. No. 0.0019;
Matches 20; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
OY      1 RWIRVQRCRAIRHPRIRROGLRMLR 29
DB      824 RVEVVOEAYRAIRHPRIRROGLERIL 852
RESULT  9
ENV_HV1BN STANDARD: PRT; 853 AA.
AC      P04581;
ID      ENV_HV1BN
DT      13-AUG-1987 (Rel. 05, Created)
DT      13-AUG-1987 (Rel. 05, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Envelope polypeptide GP160 precursor [Contains: Exterior membrane
DE      glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN      ENV.
OS      Human immunodeficiency virus type 1 (ELI isolate) (HIV-1).
OC      Viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11689;
RN      [1]
RP      SEQUENCE FROM N.A. MEDLINE=86245056; PubMed=2424612;
RX      MEDLINE=86245056; PubMed=2424612;
RA      Allison M., Main-Hobson S., Montagnier L., Sonigo P.;
RT      "Genetic variability of the AIDS virus: nucleotide sequence analysis
RT      of two isolates from African patients.";
RL      Cell 46:63-74(1986).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; K03454; AAA44329.1; -
DR      EMBL; A07108; CAA00616.1; -
DR      HIV; K03454; ENVSELI.
DR      InterPro: IPR000328; Env GP41.
DR      InterPro: IPR000777; GP120.
DR      Pfam; PF00516; GP120; 1.
DR      Pfam; PF00517; GP41; 1.
KW      Aids; Coat protein; Glycoprotein; Transmembrane;
KW      signal.
FT      SIGNAL 1 31
FT      CHAIN 32 508
FT      CHAIN 509 853
FT      DISULFID 53 73
FT      DISULFID 118 206
FT      DISULFID 125 197
FT      DISULFID 130 154
FT      DISULFID 219 248
FT      DISULFID 229 240
FT      DISULFID 297 330
FT      SIGNAL 32 31
FT      CHAIN 32 508
FT      CHAIN 509 853
FT      DISULFID 53 73
FT      DISULFID 118 206
FT      DISULFID 125 197
FT      DISULFID 130 154
FT      DISULFID 219 248
FT      DISULFID 229 240
FT      DISULFID 297 330

```

```

FT DISULFID 376 442 BY SIMILARITY.
FT DISULFID 383 416 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 853 AA; 96721 MW; P9CDB64DAAD07A5 CRC64;

```

```

Query Match 44.9% Score 79; DB 1; Length 853;
Best Local Similarity 60.7%; Pred. No. 0.0019;
Matches 17; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
Qy 1 RWIRVQRCRAIRHRRIRROGLRRWL 28
Db 825 RVIEIQRACRAVNIPIRRIRGLELRL 852

```

```

RESULT 10
ENV_HV1MF STANDARD; PRT: 853 AA.
AC P19551;
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (MFA isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11704;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90317877; PubMed=1695254;
RA Stevenson M., Haggerty S., Lamonica C., Mann A.M., Meier C.,
RA Wasiaik A.;
RT "Cloning and characterization of human immunodeficiency virus type 1
RT variants diminished in the ability to induce syncytium-independent
RT cytolysis."
RL J. Virol. 64:3792-3803(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M33943; AAA44850.1; -.
DR HIV: M33943; ENVSME.
DR InterPro: IPR000328; Env_GP41.
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120.1.
DR Pfam: PF00517; GP41.1.
KW Aids; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
Signal.
FT SIGNAL 1 30
FT CHAIN 31 509 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 510 853 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 54 74 BY SIMILARITY.
FT DISULFID 119 203 BY SIMILARITY.
FT DISULFID 126 194 BY SIMILARITY.
FT DISULFID 131 157 BY SIMILARITY.
FT DISULFID 216 245 BY SIMILARITY.
FT DISULFID 226 237 BY SIMILARITY.
FT DISULFID 294 329 BY SIMILARITY.
FT DISULFID 376 443 BY SIMILARITY.
FT DISULFID 383 416 BY SIMILARITY.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 614 614 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 853 AA; 96912 MW; 3377B993B6F22ABA CRC64;

```

```

Query Match 44.3% Score 78; DB 1; Length 853;
Best Local Similarity 71.4%; Pred. No. 0.0026;
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 1 RWIRVQRCRAIRHRRIRROGLRRWL 28
Db 826 RVIEVQAGAVRAIRHRRIRGLELRL 853

```

RESULT 11

ENV\_HV122 STANDARD; PRT: 853 AA.

AC P12487;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].

Query Match	Score	DB	Length
Matches 17: Conservative	44.3%	78	855
	60.7%	Pred. No. 0.0026	
	3	Mismatches	8
		Indels	0
		Gaps	0

[illegible]



```

FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .)
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .)
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .)
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .)
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .)
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .)
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .)
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .)
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .)
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .)
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .)
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .)
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .)
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .)
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .)
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .)
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .)
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .)
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .)
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .)
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .)
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .)
FT CARBOHYD 625 625 N-LINKED (GLCNAC. . .)
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .)
FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .)
FT CARBOHYD 750 750 N-LINKED (GLCNAC. . .)
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .)
SQ SEQUENCE 856 AA; 97224 MW; 0BFFB1A18931BB27 CRC64;

```

```

Query Match 44.3%; Score 78; DB 1; Length 856;
Best Local Similarity 71.4%; Pred. No. 0.0026;
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 1 RWIRVQRCRAIRHRIWRIQGLRRL 28
| | | | | | | | | | | | | | | |
DB 828 RVIEVQAGVRAIRHRIWRIQGLRRL 855

```

```

RESULT 13
ENV_HV1PV STANDARD; PRT; 856 AA.
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (PV22 isolate) (HIV-1).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11700;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8511157; PubMed=2982104;
RA Muesing M.A., Smith D.H., Cabralilla C.D., Benton C.V., Lasky L.A.,
RT "Nucleic acid structure and expression of the human
RT AIDS/lymphadenopathy retrovirus."
RL Nature 313:450-458(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K02083; AAB59873.1;
CC EMBL: X01762; CAA25903.1; ALT_SEQ.
CC PIR: A03974; VCLJVL.
CC HIV: K02083; ENVSPV22.

```

```

DR InterPro: IPR000328; Env_GP41.
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120.1.
DR Pfam: PF00517; GP41.1.
KW AIDS: Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 30
FT CHAIN 31 511
FT CHAIN 512 856
FT DISULFID 54 74
FT DISULFID 119 205
FT DISULFID 126 196
FT DISULFID 131 157
FT DISULFID 218 247
FT DISULFID 228 239
FT DISULFID 296 331
FT DISULFID 378 445
FT DISULFID 385 418
FT CARBOHYD 88 88
FT CARBOHYD 136 136
FT CARBOHYD 141 141
FT CARBOHYD 156 156
FT CARBOHYD 160 160
FT CARBOHYD 186 186
FT CARBOHYD 197 197
FT CARBOHYD 230 230
FT CARBOHYD 234 234
FT CARBOHYD 241 241
FT CARBOHYD 262 262
FT CARBOHYD 276 276
FT CARBOHYD 289 289
FT CARBOHYD 295 295
FT CARBOHYD 301 301
FT CARBOHYD 332 332
FT CARBOHYD 339 339
FT CARBOHYD 356 356
FT CARBOHYD 386 386
FT CARBOHYD 392 392
FT CARBOHYD 397 397
FT CARBOHYD 406 406
FT CARBOHYD 448 448
FT CARBOHYD 463 463
FT CARBOHYD 611 611
FT CARBOHYD 616 616
FT CARBOHYD 625 625
FT CARBOHYD 637 637
FT CARBOHYD 674 674
FT CARBOHYD 750 750
FT CARBOHYD 816 816
SQ SEQUENCE 856 AA; 97339 MW; 5FCDB1DC3C1209B3 CRC64;

```

```

Query Match 44.3%; Score 78; DB 1; Length 856;
Best Local Similarity 71.4%; Pred. No. 0.0026;
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 1 RWIRVQRCRAIRHRIWRIQGLRRL 28
| | | | | | | | | | | | | | | |
DB 828 RVIEVQAGVRAIRHRIWRIQGLRRL 855

```

```

RESULT 14
ENV_HV1A2 STANDARD; PRT; 855 AA.
AC P03378;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (ARV/SF2 isolate) (HIV-1).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.

```

OX NCBI\_TaxID=11685;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=855090453; PubMed=2578227;  
 RA Sanchez-Pescador R., Power M.D., Barr P.J., Steimer K.S.,  
 RA Stempelen M.M., Brown-Shimer S.L., Gee W.W., Renard A., Randolph A.,  
 RA Levy J.A., Dina D., Luciw P.A.;  
 RT "Nucleotide sequence and expression of an AIDS-associated retrovirus  
 (ARV-2).";  
 RL Science 227:484-492(1985).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K02007; AAB59882.1; -  
 DR PIR; A03976; VCLJAZ.  
 DR HIV; K02007; ENVSE2.  
 DR InterPro; IPR000328; Env\_Gp41.  
 DR InterPro; IPR000777; Gp120.  
 DR Pfam; PF00516; Gp120; 1.  
 DR Pfam; PF00517; Gp41; 1.  
 KW Aids; Coat protein; Polyprotein; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT CHAIN 1 29  
 FT SIGNAL 30 509  
 FT CHAIN 510 855  
 FT DISULFID 53 73  
 FT DISULFID 118 208  
 FT DISULFID 125 199  
 FT DISULFID 130 155  
 FT DISULFID 221 250  
 FT DISULFID 231 242  
 FT DISULFID 299 333  
 FT DISULFID 380 442  
 FT DISULFID 387 415  
 FT CARBOHYD 87 129  
 FT CARBOHYD 129 140  
 FT CARBOHYD 140 154  
 FT CARBOHYD 154 158  
 FT CARBOHYD 158 184  
 FT CARBOHYD 184 190  
 FT CARBOHYD 190 200  
 FT CARBOHYD 200 223  
 FT CARBOHYD 223 233  
 FT CARBOHYD 233 244  
 FT CARBOHYD 244 265  
 FT CARBOHYD 265 279  
 FT CARBOHYD 279 292  
 FT CARBOHYD 292 298  
 FT CARBOHYD 298 304  
 FT CARBOHYD 304 334  
 FT CARBOHYD 334 341  
 FT CARBOHYD 341 358  
 FT CARBOHYD 358 364  
 FT CARBOHYD 364 388  
 FT CARBOHYD 388 394  
 FT CARBOHYD 394 400  
 FT CARBOHYD 400 408  
 FT CARBOHYD 408 445  
 FT CARBOHYD 445 458  
 FT CARBOHYD 458 461  
 FT CARBOHYD 461 610  
 FT CARBOHYD 610 615  
 FT CARBOHYD 615 624  
 FT CARBOHYD 624 636  
 FT CARBOHYD 636 815  
 FT CARBOHYD 815 855  
 SEQUENCE 855 AA; 97438 MW; A3BC20573AAC41A2 CRC64;

Query Match 43.8%; Score 77; DB 1; Length 855;  
 Best Local Similarity 67.9%; Pred. No. 0.0036;  
 Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 Oy 1 RWIRVORNCRAIRHWRIRROGLRRL 28  
 Db 827 RVIEVAQRARVRAIRHWRIRROGLRRL 854  
 RESULT 15  
 EN\_VH1B8 STANDARD; PRT; 851 AA.  
 ID EN\_VH1B8  
 AC P04582;  
 DR 13-AUG-1987 (Rel. 05, Created)  
 DR 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Envelope glycoprotein Gp160 precursor [Contains: Exterior membrane  
 DE glycoprotein (Gp120); Transmembrane glycoprotein (Gp41)].  
 GN ENV.  
 OS Human immunodeficiency virus type 1 (BH8 isolate) (HIV-1).  
 OC Viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11684;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=85111123; PubMed=2578615;  
 RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,  
 RA Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A.,  
 RA Bauneister K., Ivanoff L., Peteway S.R., Jr., Pearson M.L.,  
 RA Lautenberger J.A., Pappas T.S., Chrayeb J., Chang N.T., Gallo R.C.,  
 RA Wong-Staal F.;  
 RT "Complete nucleotide sequence of the AIDS virus, HMLV-III.";  
 RL Nature 313:277-284(1985).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K02011; AAA44661.1; -  
 DR HIV; K02011; ENVSBH8.  
 DR Gp41; K02011; ENVSE2.  
 DR InterPro; IPR000328; Env\_Gp41.  
 DR InterPro; IPR000777; Gp120.  
 DR Pfam; PF00516; Gp120; 1.  
 DR Pfam; PF00517; Gp41; 1.  
 KW Aids; Coat protein; Polyprotein; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT CHAIN 1 30  
 FT SIGNAL 31 506  
 FT CHAIN 507 851  
 FT DISULFID 54 74  
 FT DISULFID 119 205  
 FT DISULFID 126 196  
 FT DISULFID 131 157  
 FT DISULFID 218 247  
 FT DISULFID 228 239  
 FT DISULFID 228 239  
 FT DISULFID 296 331  
 FT DISULFID 378 440  
 FT DISULFID 385 413  
 FT DISULFID 413 413  
 FT CARBOHYD 88 88  
 FT CARBOHYD 136 136  
 FT CARBOHYD 141 141  
 FT CARBOHYD 141 141  
 FT CARBOHYD 156 156  
 FT CARBOHYD 156 160  
 FT CARBOHYD 160 160  
 FT CARBOHYD 186 186  
 FT CARBOHYD 197 197  
 FT CARBOHYD 230 230  
 FT CARBOHYD 234 234  
 FT CARBOHYD 241 241  
 SEQUENCE 855 AA; 97438 MW; A3BC20573AAC41A2 CRC64;



